A NEURAL NETWORK BASED WORKSTATION FOR AUTOMATED CELL PROLIFERATION ANALYSIS

F. Arámbula Cosío[1], L. Vega [2], A. Herrera Becerra[1], R. Prieto Meléndez[1], G. Corkidi.[2]
[1]Centro de Instrumentos, UNAM. P.O. Box 70-186, México 04510, D.F.
[2]Instituto de Biotecnología, UNAM. P.O. Box 510-3, 62250, Cuernavaca, Morelos, México

Abstract-**In this paper is reported the development of a neural network (NN) based workstation for automated cell proliferation analysis, of cytological microscope images. The software of the system assists the expert biotechnologist during cell proliferation and chromosome aberration studies by automatically identifying metaphase spreads and stimulated nuclei on each digital image. After manual edition of metaphase false positives, the system automatically calculates the mitotic index (MI) i.e. the ratio of metaphases to stimulated nuclei of a given tissue sample. The system reported has been able to classify correctly approximately 91% of the metaphases and stimulated nuclei, in a test set of 191 mitosis, 331 nuclei, and 387 artefacts, obtained from 30 different microscope slides. Manual edition of false positives from the metaphase classification results allows the calculation of the MI with an error of 6.5%.**
Keywords - **automated object recognition, mitotic index, metaphase finder**

I. INTRODUCTION

Modern development of a variety of chemical products used in industry, pharmaceuticals, cosmetics, and food additives, has created the need for fast and effective methods to evaluate its effects on cellular proliferation [1]. A reliable endpoint to evaluate and compare cell proliferation rates is the mitotic index (MI), which is the percentage of cells that are in the process of division. The mitotic index is usually determined through light-microscope analysis of slide preparations. The analyst identifies at least 2000 cells per slide and calculates the percentage of metaphase spreads found among the interphase or "stimulated" nuclei. Metaphase identification on microscope slides is also performed during the scoring of radiation-induced chromosomal aberrations. This scoring is performed in order to assess the effects of radiation exposure due to medical treatment, accidental, or environmental exposure. This type of procedure is also labour intensive. For example, in order to detect exposure to low radiation doses of X or gamma rays, the frequency of occurrence of diccentric chromosomes in 1000 metaphases must be analysed [2].

Previous work on automatic metaphase finders includes: the Genetiscanner, with a true positive rate of 80% and a false positive rate of 20% [3]. Reference [4] reports a supervised size and circularity criterion to detect metaphases, which provides a 78% true positive rate. Reference [5] reports an automatic system for metaphase identification and chromosome aberration analysis on preparations stained with flourescent dyes, a true positive rate (during metaphase identification) of 87.3%, and a false positive rate of 7.4%. Reference [6] reports a texture feature to classify previously segmented objects, into metaphase spreads and interphase nuclei, with true positive rates of 84% and 87% respectively. Reference [2] reports a system for automatic metaphase identification using a second derivative feature to detect the chromosomes inside of a metaphase. The true positive rate of the system is 74% with a false positive rate of 6 %.

In this paper is presented a NN-based workstation for improved automatic identification of metaphase spreads and nuclei on microscope slide images. Each microscope slide is automatically scanned for each of the fields of the microscope. Image processing techniques are used to segment the objects on each image. Ten different morphological features are measured on each segmented object. A neural net is used to classify each ten-feature vector into metaphase spreads and stimulated nuclei. Providing in this way automatic metaphase and nuclei identification during MI calculation, as well as automatic metaphase identification for manual chromosome aberration analysis. Given the small ratio of metaphases to nuclei involved during MI calculation, manual deletion of false positives from the metaphases annotated by the system is necessary.

II. SYSTEM DESCRIPTION

The image acquisition system consists of an optical microscope (Olympus BH2) with a motorised plate (Märzhauser, Germany) and a CCD B&W video camera attached (Cohu 4800). A 10X objective lens is used during image acquisition. A Matrox frame grabber with a 512x480 pixel resolution was used for digitisation. The sample preparation details are described in [6].

A. Image Segmentation

The object types for automated cell proliferation study purposes are: Metaphases (M), which include compact metaphase spreads (CM), and scattered metaphase spreads (SM); Stimulated nuclei (SN); and Artefacts (AF), which include non-stimulated nuclei (NSN) and cellular debris (CD). Examples of each object are shown in Fig 1. Digital images are pre-processed with a fourth order function to enhance the contrast of the stimulated nuclei [6].

Recursive dilation [7] is next applied to each digital image to join the chromosomes inside scattered metaphase spreads. Pre-processed images are segmented by minimisation of within group variance [8]. The segmentation process

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
25 Oct 2001	N/A	-

Title and Subtitle A Neural Network Based Workstation for Automated Cell Proliferation Analysis	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Centro de Instrumentos UNAM. P.O.Box 70-186 Mexico 04510, D.F.	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research, Development & Standardization Group (UK) PSC 802 Box 15 FPO AE 09499-1500	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes
Papers from 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, October 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom.

Abstract

Subject Terms

Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU

Number of Pages
4

annotates each object on the digital image with a one-pixel width, closed, contour, as shown in Fig.1. Large artefacts and non-stimulated nuclei are eliminated at this stage applying an outlier exclusion criterion as follows. All objects with an area outside of the range [mean nuclei area - 3σ, mean metaphase area + 3σ] are discarded. This procedure eliminates an average of 58% of the artefacts on each image.

III. Feature extraction

Ten morphological features were used to characterise each segmented object. The approach followed to select appropriate object features, was to calculate features similar to those used by a human expert, during image annotation.

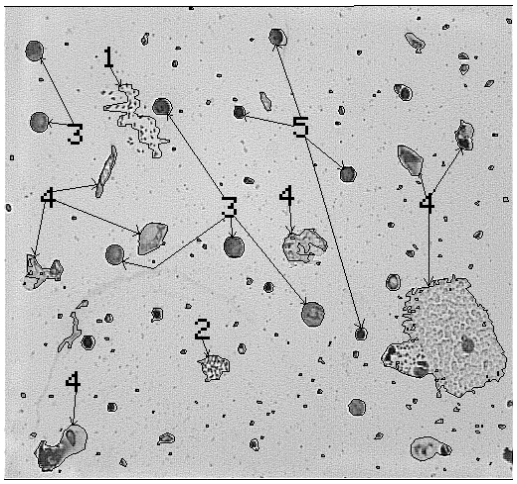

Fig. 1. Segmented microscope image. 1. (SM) Scattered metaphase spread, 2. (CM) Conglomerated metaphase spread, 3. (SN) Stimulated nuclei, 4. (CD) Cellular debris, 5. (NSN) Non-stimulated nuclei.

A. Nuclei Identification Features

The human expert identifies stimulated nuclei mainly by its circular shape, grey level characteristics, and size. In consequence, the following features were included in the feature vectors of each object: form factor (FF), grey level mean (GM) and standard deviation (GSD), area (A).

B. Metaphase identification features

The human expert identifies metaphase spreads mainly by the internal texture produced by the chromosomes inside each metaphase. Thus in order to increase the percentage of true positives, and to decrease the percentage of false positives during metaphase and nuclei classification, 5 textural features were added to the feature vectors of each object.

The MDWRE [6] is the mean value of the depth-width ratio of the troughs in a horizontal scan line of the object image (Fig. 2). The standard deviation of the MDWRE (MDWRESD) of each object was included in the feature vectors, in order to detect the heterogeneity in the depth-width ratios of the troughs of a given object.

Three parameters related with the relative extrema density [9] were included to detect, at different scales, textural features due to the chromosomes inside of a metaphase. We defined the absolute extrema density (AED) as the number of crossings of a certain threshold value, on a horizontal scan-line of the object image as shown in Fig. 2. The threshold value in Fig. 2 is calculated using the method described in [8], which corresponds to the optimum grey level value to segment chromosomes from the background on a metaphase image. Horizontal scan lines on each object image were sampled every 4 and every 20 pixels in the vertical direction. On each scan line, crossings were measured as shown in Fig.2. The total number of crossings at each line-sampling value were normalised dividing by the total object area in pixels These measures were named NC4/area and NC20/area respectively. An average measure of texture was calculated as the total number of crossings, counted in all image lines per object area. This measure was named NC/area.

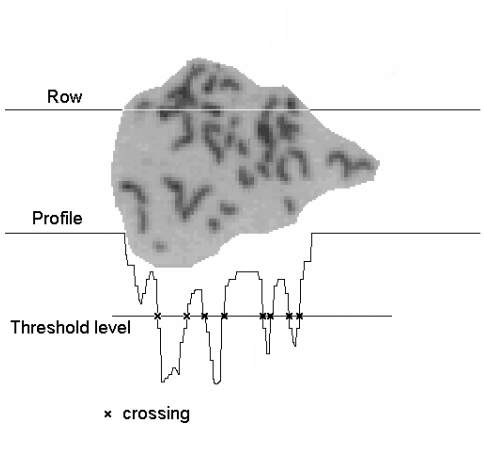

Fig. 2 Absolute extrema density measurement

Cumulative grey level histograms of scattered metaphases showed a characteristic slope change as shown in Fig. 3. This is because scattered metaphases have a significant amount of homogeneous clear background (P1-P2 region), with dark stains corresponding to the chromosomes (P0-P1 region). P1 corresponds to the BCV (Between Class Variance) grey level [8]. This grey level value is the optimal threshold separating background from chromosomes. Since P0, P1 and P2 are located just in the knees of the cumulated histogram, we defined intermediate points P0', P1', P1" and P2' in order to characterise the line segments (P0'-P1') and (P1"-P2'), where: P0'=P0+0.1*P2; P1'=0.9 *P1; P1''=1.1*P1; P2'=0.9*P2. The histogram slope difference (CHSD) was calculated as the absolute difference of slopes of the first (P0'-P1') and the second (P1''-P2') histogram sections, as shown in Fig.3. The measure was normalised dividing by the total object area (CHSD/area).

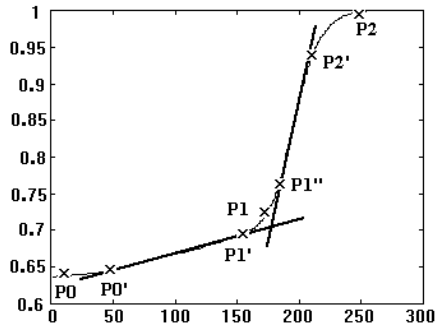

Fig. 3, Histogram slope difference calculation

IV. NEURAL NET CONSTRUCTION AND EVALUATION

A three layer feedforward architecture was used in this work for the different neural nets implemented for metaphase, nuclei, and artefact classification [10]. A data set of 909 patterns - 191 metaphases, 331 nuclei, and 387 artefacts - taken from 30 different microscope slides, was used to train and test each different NN. Each pattern included the ten features described in section III. The training data consisted of 80 metaphases, 135 nuclei and 150 artefacts, taken at random from the data set. The evaluation set consisted of the remaining non-training patterns in the data set - 111 metaphases, 196 nuclei, and 237 artefacts. All NNs were trained using backpropagation with momentum (0.95) and adaptive learning rate (initial value of 0.01, learning rate increase of 1.05, learning rate decrease of 0.7, and maximum error ratio of 1.04). The hidden units use a hyperbolic tangent as activation function, and the output units use the logistic function [11].

Since each NN output can take any values between 0 and 1, we followed an error criteria in order to assign a pattern to a certain class. The usual approach is to calculate the mean square error (4) and assign the pattern to the class with maximum output if the error is smaller than a selected threshold value. In this work a threshold value of 0.05 was used, this value minimises the number of misclassified objects. If the output error (4) is larger than 0.05 the corresponding input pattern is counted as a non-classified.

$$e = \sqrt{\sum_{i=1}^{i=k} [O_i - b]^2}\,; b = \begin{Bmatrix} 0.9 \; if \;\; Oi = MAX \\ 0.1 \quad otherwise \end{Bmatrix} \qquad (4)$$

where
k is the number of classes (k=3 in this case);
Oi is the output unit i of the neural net.

Twelve NNs were constructed, each with 10 input units, a varying number of hidden units (between 2 and 15), and three output units one for each class. In order to select the best performing NN (optimal number of hidden units), we have to take into account the different miss-classification errors produced by the NN, emphasizing those errors that are most costly. These errors are typically specified in a confusion matrix [12]. Since we developed three-class classifiers we have a 3-by-3 confusion matrix for each NN, with 3 correct classifications and 6 different errors the classifier can make. Additionally we have a certain number of non-classified objects (i.e. the objects that the NNs were not able to include in any of the specified three classes).

In order to consider all the terms in the confusion matrix plus the proportion of non-classified objects, an *adhoc* performance measure has been constructed. We have considered the difference of all correctly classified patterns (true positives) minus the sum of the lost (false negatives) and misclassified (false positives) patterns multiplied by a weight (i.e. error cost value). The best NN is the one that maximises (5). A perfect NN would have a value of 1 (for 100% efficacy), and a totally imperfect one would have a value of -1 (for 0% efficacy).

$$NN_{pfm} = \sum_{i \in \{M,SN,AR\}} \left[\frac{\rho_i}{K_i} \left(C_{ii} - \sum_{\substack{j \neq i \\ j \in \{M,SN,AR\}}} [C_{ij} + C_{ji}] - NCL_i \right) \right] \qquad (5)$$

where
ρ_i is the weight (i.e. error cost value) of the i class;
Ki is the number of patterns assigned to each of the classes metaphases, nuclei, and artefacts;
Cij are the elements of the confusion matrix;
NCLi is the number of non-classified objects of each class.

The mitotic index is defined as the ratio of metaphases (scattered + conglomerated) to stimulated nuclei as shown in (3). Usually 2000 objects (metaphases + stimulated nuclei) are used in the calculation of the MI

$$MI = N_M / N_{SN} \qquad (3)$$

where N indicates the number of objects of the classes metaphases (M) and stimulated nuclei (SN).

Typical MI values are between 2% to 5% for 2000 objects counted (M + SN = 2000). We assigned the following weights (ρ_i) for a mean MI of 3,5%: ρ_M = 0,965, ρ_n = 0,035, ρ_{AR} = 0.0. We have assigned a value of zero to this last weight since artefacts AR are not involved during MI calculation. Artefacts is a class created for a better identification of the two relevant classes M and SN. In other words, we don't mind if artefacts are for example non-classified or well classified, but we care if they are assigned erroneously to the other classes (in which case ρ_M and ρ_n take this into account). The two best performing NNs were the 10-9-3 (NN_{pfm}= 0.776) and the 10-15-3 (NN_{pfm}= 0.777). Table I shows a comparison of the confusion matrices for these NNs. Table II shows the proportion of non-classified objects.

TABLE I
CONFUSION MATRICES FOR THE 10-9-3 AND 10-15-3 NNS

	M		N		AR	
	10-9-3	10-15-3	10-9-3	10-15-3	10-9-3	10-15-3
M	0.918	0.918	0	0	0.054	0.027
SN	0	0	0.918	0.939	0.046	0.030
AR	0.029	0.029	0.025	0.034	0.894	0.848

TABLE II
NON-CLASSIFIED OBJECTS

	Non-Classified	
	10-9-3	10-15-3
M	0.027	0.054
SN	0.036	0.031
AR	0.051	0.088

V. DISCUSSION

Table III shows, for the best two selected NNs, the expected numbers of metaphases, nuclei and artefacts for an MI of 3.5%. The expected sample sizes would be: 68 metaphases, 1932 stimulated nuclei, 4900 artefacts.

TABLE III
EXPECTED NUMBERS OF CELLS AND ARTEFACTS FOR 10-9-3 AND 10-15-3 NNS FOR MI CALCULATION (MI=3.5%). ALL VALUES ARE IN NO. OF OBJECTS

	M		SN		AR	
	10-9-3	10-15-3	10-9-3	10-15-3	10-9-3	10-15-3
M	62	62	0	0	4	2
SN	0	0	1774	1814	89	59
AR	145	145	124	165	4383	4156
Total	207	207	1898	1979	4476	4217

As we can see in Table III the number of metaphase false positives (145 for both NNs) is small enough for manual selection of true metaphases. At the last stage of analysis, our instrument displays to the user a final screen containing the shapes of all objects classified as metaphases. The user invests around two additional minutes to select with a pointer the true metaphases, this is a negligible amount of time compared to the 40 hours needed for completely manual MI calculation. With this simple user intervention, the overall accuracy of the instrument increases to 6.47% for the 10-9-3 NN.

VI. CONCLUSIONS

The development of an automated system for cell proliferation analysis has been presented. A neural net classifier is used for semi-automatic MI calculation during cell proliferation studies as well as for chromosome aberration analysis, providing automatic identification of metaphase spreads and nuclei. The use of 10 morphometrical, photometrical, and textural features to train neural networks for automatic recognition of metaphases, nuclei, and artefacts in microscope images at low magnification values (10X), has been reported. Low magnification values enable a fast scanning of the microscope slides. The best performing neural net classifier (10-9-3) has been able to provide false negative, and false positive rates, suitable for practical use during automatic identification of metaphases, outperforming all previously reported systems for automatic identification of metaphases.

The system reported here used in conjunction with a systematic (i.e. repeatable) preparation of tissue samples [6], has the potential to achieve a performance suitable for regular laboratory use during automatic identification of metaphases and semi-automatic MI calculation in microscope images at low (10X) magnification.

ACKNOWLEDGMENT

This work was supported by the *Programa de Apoyo a Proyectos de Desarrollo e Investigacíon en Informática REDII 2000*. We thank Blanca Itzel Taboada for her technical assistance.

REFERENCES

[1] F.A. Barile , (Ed.), "*In vitro* cytotoxicology. Mechanisms and methods", *CRC Press*, USA, pp. 1-222, 1994.

[2] J.R.N. McLean and F. Johnson, "Evaluation of a metaphase chromosome finder: Potential Application to chromosome-based radiation dosimetry", Micron, vol.26, No. 6, pp. 489-492, 1995.

[3] K.R. Castleman, "The PSI automatic metaphase finder", *J. Radiat. Res.*, 33, (Supl.), pp. 124-128, 1992.

[4] M. Garza-Jinich, C. Rodriguez, G. Corkidi, R. Montero, E. Rojas, P. Ostrosky-Wegman, "A microcomputer-based supervised system for automatic scoring of mitotic index in citotoxicity studies", in Archibald C., and Petriu E., (Eds.), "Advances in machine vision, vol. 32", *World Scientific Press series on Computer Science*, pp. 301-313, 1992.

[5] J. Vrolijk, W.C. Sloos, F. Darroudi, A.T. Natarajan, and H.J. Tanke, "A system for flourescence metaphase finding and scoring of chromosomal translocations visualized by in situ hybridization", *Int. J. Radiat. Biol.*, 66, (3), pp. 287-295, 1994.

[6] G. Corkidi, L. Vega, J. Márquez, E. Rojas, P. Ostrosky-Wegman, "Roughness feature of metaphase chromosome spreads and nuclei for automated cell proliferation analysis", *Med. Biol. Eng. Comput.*, vol 36, No 6, pp. 679-685, 1998.

[7] J. Serra, "Image analysis and mathematical morphology", *Academic Press*, pp. 43-49, 1989.

[8] N. Otsu, "A threshold selection method from gray-level histograms", *IEEE Transactions on Systems Man and Cybernetics*, vol. 9, No. 1, pp. 62-66, 1979.

[9] A. Rosenfeld, and E Troy, "Visual texture analysis", Technical report, *University of Maryland*, College Park, Maryland, USA, pp.70-116, 1970.

[10] F. Arámbula Cosío, L. Vega, A. Herrera Becerra, R. Prieto Meléndez, G. Corkidi, "Automatic identification of metaphase spreads and nuclei using neural networks", Accepted in *Med. Biol. Eng. Comput.*, 2001.

[11] J. Hertz, A. Krog, R.G. Palmer, "Introduction to the theory of neural computation", Lecture Notes, Santa Fe Instute, vol. 1, *Addison-Wesley*, 1991.

[12] K.R. Castleman, and B.S. White, "Dot count proportion estimation in FISH specimens", *Bioimaging*, vol. 3, pp.88-93, 1995.